This is a working document prepared by the Energy Information Administration (EIA) in order to solicit advice and comment on statistical matter from the American Statistical Association Committee on Energy Statistics. This topic will be discussed at EIA's spring 2006, meeting with the Committee to be held April 6 and 7, 2006.

Version: April 3, 2006 a.m.

An Empirical Evaluation of the Relationship Between Crude Oil and Natural Gas Prices

Jose Villar jose.villar@eia.doe.gov (202)586-9613

Economic theory suggests that natural gas and crude oil prices should be related because natural gas and crude oil are substitutes in consumption and also complements, as well as rivals, in production. In general, the pattern of crude oil and natural gas prices tend to support this observation (Figure 1). However, there have been periods in which natural gas and crude oil prices have appeared to move independently of each other. Furthermore, over the past 5 years, periods when natural gas prices have appeared to decouple from crude oil prices have been occurring with increasing frequency with natural gas prices rising above its historical relationship with crude oil prices in 2001, 2003, and again in 2005. This has led some to conclude that natural gas and crude oil prices are no longer related.

Economic factors link oil and gas prices through both supply and demand. Economic theory and market behavior suggest that past changes in the oil price drive changes in the natural gas price, but the converse does not appear likely. One reason for the asymmetric relationship is the relative size of each market. The crude oil price is determined on the world market, while natural gas markets tend to be regionally segmented. Consequently, the domestic natural gas market is much smaller than the global crude oil market, and events or conditions in the U.S. natural gas market seem unlikely to be able to influence the global price of oil.

This paper seeks to develop an understanding the salient characteristics of the economic and statistical relationship between oil and gas prices. The analysis identifies the economic factors suggesting how crude oil and natural gas prices are related, and assesses the statistical significance of the relationship between the two over time. A significant stable relationship between the two price series is identified. Oil prices are found to influence the long run development of natural gas prices, but are not influenced by them.

Economic Factors Linking Natural Gas and Crude Oil Prices

Increases in oil prices may affect the natural gas market in a several ways.

Demand:

- *An increase in crude oil prices motivates consumers to substitute natural gas for petroleum products in consumption, which increases natural gas demand and hence prices*. Oil and natural gas are competitive substitutes primarily in the electric generation and industrial sectors of the natural gas market. The National Petroleum Council (NPC) in its 2003 report estimated that approximately 5 percent of industrial boilers can switch between natural gas and petroleum fuels. Other analysts estimate that up to 20 percent of power generation capacity is dual-fired, although in practice it is expected that the relevant percentage is considerably less. However, fuel-switching is not limited to dual-fired units. Some degree of additional fuel switching is achieved by dispatching decisions to switch from single-fired boilers of one type to that of another fuel. Although these are limited percentages, the shift in marginal consumption can have a pronounced impact on prices in a tight market.

Supply:

- *Increases in crude oil prices may increase natural gas produced as a co-product of oil, which would tend to decrease natural gas prices*. Natural gas is found in two basic forms—associated gas and non-associated gas. Associated gas is natural gas that occurs in crude oil reservoirs either as free gas (associated) or as gas in solution with crude oil (dissolved gas). Non-associated gas is natural gas that is not in contact with significant quantities of crude oil in the reservoir. In 2004, associated-dissolved gas comprised approximately 2.7 trillion cubic feet (Tcf) or 14 percent of marketed natural gas production in the United States.
- *An increase in crude oil prices may lead to increased costs of natural gas production and development, putting upward pressure on gas prices*. Natural gas and crude oil operators compete for similar economic resources such as labor and drilling rigs. An increase in the price of oil would lead to higher levels of drilling or production activities as operators explored for and developed oil prospects at a higher rate. The increased activity would bid up the cost of the relevant factors, which will increase the cost of finding and developing natural gas prospects.
- *An increase in crude oil prices may lead to more drilling and development of natural gas projects, which would tend to decrease gas prices*. Increased oil prices affect the cash flow available to finance new drilling and project development. Changes in the relative price structure could lead to increases in drilling for one fuel at the expense of the other. However, it is generally expected that increased cash flow would expand supply activities for both gas and oil.

These economic factors suggest that oil and gas prices should be related. The analysis of crude oil and natural gas prices that follows empirically tests this hypothesis by drawing on the extensive time series literature on nonstationary processes, and cointegration. In general, it should not be possible to form a weighted average of two nonstationary variables and create a stationary time series. However, in certain cases when two variables share common stochastic or deterministic trends, it is possible to create such a cointegrating relationship. Because cointegrated variable shares an intrinsic structural relationship over time, empirically testing for the presence of cointegration constitutes an empirical test of the long-run relationship between the variables.

The analysis will proceed as follows. First, the key time series concepts used in the analysis will be presented, focusing on the properties of nonstationary variables and how they relate to the notion of cointegration. Second, the time series properties of natural gas and crude oil prices will be examined, emphasizing the data generating process that created them. Both price series will be identified as nonstationary variables indicating the applicability of cointegration analysis. Next, the vector autoregression model will be introduced as a method to summarize multivariate time series, emphasizing its relationship to the vector equilibrium correction model and cointegration. Finally the model will be estimated and hypothesis testing will be done. Natural gas and crude oil prices are identified as cointegrated variables, and the implications of the findings will be discussed.

Key Time Series Concepts

Characteristics of Stationary and Nonstationary Time Series

Econometric analysis of the classical linear regression model depends heavily on the assumption that observed data result from stationary data generation processes. However, in practice many economic time series are nonstationary. Stationary time series evolve independently of time.[1] Specifically, stationarity requires that random shocks or innovations tend to dissipate rapidly and not to have any lasting effects on the evolution of the time series. In contrast, non-stationary time series have sustained effects resulting from random shocks. For example, consider the following autoregressive process:

(1) $y_t = \rho y_{t-1} + u_t$,

where u_t is assumed to be a purely random variable that is normally, independently, and identically distributed with mean μ, and variance σ^2.

Depending on the value of ρ in equation 1, y_t may be either stationary or nonstationary. If $|\rho|$ is less than 1, it can be shown from equation 1 that:

(2) $y_t = \rho^t y_0 + \Sigma_i \rho^i u_{t-i.}$

Equation 2 indicates that y_t is related to sum of all random past shocks and an initial starting condition, y_0. When ρ is less than 1, random shocks to the system dissipate with time, and y_t is a stationary process.[2] This occurs because ρ^i approaches zero asymptotically as i increases. Because the impacts of random shocks to the system do not persist, it can be shown that the mean, variance, and autocovariance asymptotically converge to constants as t becomes larger. Consequently, stationary processes evolve independently of time.

If ρ=1, it can be shown that:

(3) $y_t = y_0 + \Sigma_i u_{t-i.}$

When ρ=1, y_t is equal to sum of all past shocks and an initial starting condition, y_0, such that all past disturbances have a permanent effect on y_t. In other words, a random shock that is far-removed from the current period will have as much impact as on the current

[1] Stationary time series are defined to have constant mean and variance, while the autocovariance between any two values from the series depends only on the time interval between the two values and not on the point in time.
[2] When ρ is greater than 1, past shocks become more influential over time. This case lacks practical applications in economics, and is not of interest for the current analysis.

period as a random shock of equal magnitude in the current period. As a result, it can be shown that the variance and autocovariance vary with time, depending on the initial starting conditions. It is the time-varying nature of the variance and autocovariance that generally cause the failure of the assumptions underlying the classical linear regression model: no autocorrelation, homoskedasticity, and multivariate normality. A fundamental challenge in the analysis of time series is the nonstationarity of many economic variables. For example, one well known issue associated with nonstationary variables is the problem of spurious regressions, in which a regression of two unrelated nonstationary variables--each trending upward over time—may result in a high R^2 suggesting a tight fit of the data when the model in fact explains nothing but the similar rising trend over time.

Stochastic processes such as in equation 3 are called unit roots, because $\rho=1$. A nonstationary time series such as in equation 3 is fundamentally evolutionary in nature. Starting from the initial point, y_0, y_t will either increase or decrease depending on the random outcome of u_t in a given period, accumulating each of the random shocks in each period. Consequently nonstationary time series tend to exhibit "wandering" behavior as they follow a stochastic trend—a trend with random increments. This contrasts with a deterministic trend that has constant increments.

At any given point in time, a particular observation of a nonstationary time series essentially results from a summation of all the past random errors that preceded it. This process is called integration because the present outcome incorporates all of the random errors in the past. A simple way of solving the problem of integration in a given time series is simply to subtract the value of the preceding period from the current period, which has the effect of canceling out the stochastic trends in the successive periods, and creating a stationary series. This transformation is called differencing. From equation 3, a stationary time series can be created from a nonstationary series such as y_t, by focusing on the change between the periods.

(4) $\Delta y_t = y_t - y_{t-1} = u_{t,,}$

All nonstationary variables can be de-trended by differencing. However, some nonstationary variables must be differenced more than once—taking differences of differences—to create a stationary series. The order of integration refers to the number of times that the differencing operation must be performed to restore stationarity. For example, y_t from equation 3, is integrated of order 1, or I(1).

While all I(1) variables can be differenced to generate a stationary time series, there are some I (1) variables that are cointegrated. This means that it is possible to form a linear stationary linear combination from two I(1) variables. In other words, given two variables x_t and y_t that are nonstationary I(1) variables: x_t and y_t are cointegrated if it is possible to form a linear combination such as:

(5) $y_t - \beta_0 - \beta_0 x_t = u_{t,}$ where u_t is a stationary I(0) variable

Cointegration is similar to integration in that the process of detrending the nonstationary series involves focuses on the offsetting of the shared attributes of stochastic trends. In some economic applications it is advantageous to compare two or more time series for shared stochastic patterns over time. One advantage of cointegration is that it solves the problem of spurious regressions without having to difference the data and losing the information about the levels of the time series. Cointegration of economic time series implies that the economic variables have a long-run structural relationship that can be empirically evaluated.

While the econometrics of cointegration may seem somewhat esoteric, the idea behind it is fairly simple. Imagine a man walking a dog on a leash. The dog may wander this way or that, and the man may wander this way or the other, but because the leash tethers them together, they tend to move along together. The leash prevents them from straying too far apart, and helps them gradually close the gap between them when they do separate.

The Stochastic Characteristics of West Texas Intermediate WTI Crude Oil and Henry Hub Natural Gas Prices

Using 16 years of monthly data drawn from January 1989 through December 2005, time series of WTI crude oil and Henry Hub natural gas spot prices were analyzed, focusing on decomposing the short-run and long-run effects of changes in crude oil and natural gas prices. The time period of the analysis was determined by the availability of a complete series of both Henry Hub an WTI prices. Monthly series were used because they have sufficient texture to capture short-run movements over time, without adding unnecessary complexity to the analysis.

The analysis of the time series properties of oil and gas prices begins by examining the levels of natural gas and oil prices. The natural logarithms of the WTI crude oil and Henry Hub natural gas prices are presented in the upper panel of Figure 2. The logarithmic transformations were used to remove the scale effects in the variables and reduce the possible effect of heteroskedasticity. Logarithmic transformations also have the beneficial property of allowing estimated parameters to be interpreted as elasticities.[3] While it is very difficult to tell if a time series is nonstationary by visual examination of level data, it appears that natural gas and crude oil prices do have characteristics consistent with non-stationary data. Both natural gas and crude oil prices have a meandering quality, with prices wandering up and then down. The means of the series do not appear constant over time, and in fact may be drifting upward, suggesting the possibility of a stochastic trend. The price series also appear to be serially correlated, with prices roving up for extended successive months, followed by periods of successive declines. In the lower panel, crude oil prices are shifted down gas to facilitate the a more direct comparison of the two series. The variation of the price series in the lower panel seems consistent with the possibility that natural gas and crude oil share a common trend,

[3] Throughout the remainder of the paper, the analysis will be conducted on logarithms of natural gas and crude oil prices. For the sake of brevity, the terms price and logarithm of price will be used interchangeably unless otherwise noted.

fluctuating around a fixed level. These results suggest that natural gas and crude oil prices may be cointegrated.

Histograms summarize the frequency of occurrence that a variable falls within a certain range of values. As such, histograms provide insights into the mean, standard error, and probability distribution of a given variable. Histograms of WTI and Henry Hub prices are presented in the top panel of Figure 3, with a graph of a normal distribution superimposed over the histogram. It is readily apparent that Henry Hub and WTI prices are not normally distributed. The probability distributions for both price series appear bimodal, skewed relative to the normal distribution.

The autocorrelograms of WTI and Henry Hub prices are presented in the lower panel of the figure. An autocorrelogram displays the autocorrelations of values displaced between 1 and 12 periods back. Autocorrelation appears to be significant feature of both WTI and Henry Hub prices with autocorrelations well in excess of two standard deviations from zero. For both price series, the autocorrelations decay slowly, exhibiting considerable persistence. These findings are consistent with unit root processes, which have all autocorrelations close to one. Unit root tests were conducted on both Henry Hub and WTI price series, confirming that both series are unit roots, establishing an important necessary condition for cointegration analysis.[4]

Histograms and autocorrelograms are presented Figure 4 for the first differences of both price series. In each case, the differencing transformation appears to have resolved the nonstationarities in the two series for the most part. In both cases, the probability distributions of the differenced series appear to be approximately normally distributed. Further, the autocorrelations are also much smaller, and statistically insignificant at most lags. The foregoing discussion illustrates the properties associated with nonstationary data, which can lead to increasing variances over time, autocorrelation and a failure of the assumption of multivariate normality that underlies hypothesis testing in many practical applications. These problems associated with nonstationary data will be a key consideration in the model specification and estimation that follows.

The Vector Autoregression (VAR) Model and Its Relationship to Cointegration

The vector autoregression (VAR) model provides a convenient format with which to conduct a multivariate analysis because it is able to accommodate a number of endogenous variables and lags of data. The key characteristic of a VAR is that each endogenous variable in the system is modeled on lagged values of itself and the other endogenous variables in the system. This makes the VAR very flexible, as it can facilitate the analysis of the system dynamics of random variables. Finally, the VAR can be used to identify the existence and effects of cointegration on various time series, which is inherently a multivariate phenomenon.

[4] A discussion of unit root tests and the results are presented in the Technical Appendix.

For example, the second order VAR(2) with two variables and two lags can be written:[5]

(6) $\mathbf{p}_t = \mathbf{\Pi}_1 \mathbf{p}_{t-1} + \mathbf{\Pi}_2 \mathbf{p}_{t-2} + \mathbf{B}\mathbf{x}_t + \boldsymbol{\varepsilon}_t$, where

$\boldsymbol{\varepsilon}_t \sim IN(\mathbf{0}, \mathbf{\Omega})$ is a (2x1) vector of independently, normally distributed error terms with mean 0 and symmetric covariance matrix Ω.

$\mathbf{p}_t$ is (2x1) vector of price variables.

$\mathbf{x}_t$ is a (2x1) vector is deterministic or exogenous variables.

Cointegration of economic time series implies that the economic variables have a long-run structural relationship that can be empirically evaluated. According to the Granger representation theorem, the existence of cointegration implies that an equilibrium-correction model (ECM) exists. The ECM makes it possible to distinguish between the short-run and long-run relationship between the two time series, and investigate these relationships empirically. It can be shown that equation 6 implies:

(7) $\Delta\mathbf{p}_t = \mathbf{\Gamma}_1 \Delta\mathbf{p}_{t-1} + \mathbf{\Gamma}_2 \Delta\mathbf{p}_{t-2} + \mathbf{\Pi}\mathbf{p}_{t-1} + \mathbf{B}\mathbf{x}_t + \boldsymbol{\varepsilon}_t$, where

where

$\mathbf{\Pi} = \mathbf{\Pi}_1 + \mathbf{\Pi}_2$

$\mathbf{\Gamma}_i = -\Sigma_{j=i+1} \mathbf{\Pi}_j$

The VECM representation in Equation 7 differs from the VAR in two ways. First, the VECM is expressed in both lagged levels and differences. This gives the model rich dynamics, as it incorporates both short-run effects given by the $\mathbf{\Gamma}_i$ coefficients, and the long-run effects, which are represented by $\mathbf{\Pi}$.

Second, Equation 7 explicitly contains and accounts for the effects of the cointegration vector in the expression $\mathbf{\Pi}\mathbf{p}_{t-1}$, which is the matrix algebra representation of the left-hand side of equation 5. If the equation is cointegrated, then $\mathbf{\Pi}$ may be written $\alpha\beta'$.[6]

(8) $\Delta\mathbf{p}_t = \mathbf{\Gamma}_1 \Delta\mathbf{p}_{t-1} + \mathbf{\Gamma}_1 \Delta\mathbf{p}_{t-2} + \boldsymbol{\alpha\beta}'\mathbf{p}_{t-1} + \mathbf{B}\mathbf{x}_t + \boldsymbol{\varepsilon}_t$

In this specification the long-run effects are decomposed into a speed of adjustment term, α, and the cointegrating vector β, which characterizes the long-run relationship between the levels of the variable in the model. When, the long-run relationship is in equilibrium, it has an expected value of 0. If the equilibrium is disrupted, the speed of adjustment parameter, α, determines how quickly the equilibrium is restored.

[5] Matrices and vectors are expressed in bold print.
[6] The variables are cointegrated if the rank($\mathbf{\Pi}$)>0 and less than the number of endogenous variables.

This underscores one of the major benefits of using the VECM representation. When the model is estimated in both levels and short-run changes, it does not entail the loss of information that would have occurred if a model expressed solely in differences.

Model Specification and Estimation

A general to specific approach was employed in the estimating the model. First the lag structure of the VAR is determined. While the VAR presented in Equation 6 was in 2 lags, the VAR model is general enough to accommodate any number of lag structures, so this must be determined prior to conducting the analysis. Next, a simple unrestricted version of the VAR was estimated and evaluated for fit and stability. Based on these results, the model was refined and re-estimated. Once a correctly specified model emerges from this process, restrictions are imposed on the model, and hypothesis testing is done, and the model results are interpreted.

VAR models with up to 6 lags were evaluated using the Akaike Information Criterion (AIC), and the Schwartz Bayesian Criterion (SBC). When minimized, these statistics provide guidance on which model lag structure will best fit the data. The lag structure of the VAR was determined using the AIC and SBC which indicated that a lag of two would yield the best fitting model, while maintaining a parsimonious model.

The VAR(2) of equation 6 was estimated using ordinary least squares, and mis-specification tests were performed.[7] The fitted model results and residuals are presented in Figure 5. Examination of the residuals indicates that a number of residuals appear quite large—more than three standard deviations away from the actual observations. While autocorrelation does not appear to be a significant, however it is statistically significant at the tenth and eleventh lags (Figure 6). Owing to the large residuals noted earlier with too many observations appear in the tails of the distribution, the distributions of the residuals may violate the assumption of normality. This is confirmed by examining the Jarque-Berra test statistic for normality, which with a value of nearly 19 for natural gas and 20 for crude oil, provided conclusive evidence of non-normality likely arising from leptokurtosis. The estimates are non-normal and possibly heteroskedastic.

To improve the fit of the model, several exogenous variables were added. Because natural gas demand is highly sensitive to cold weather, heating degree days (HDD) were added to the model to capture the effect of heightened demand for space heating. Next a series of centered monthly dummy variables were added to capture the seasonality of the markets. Finally, to capture the effects of tight supplies, a lagged value of the absolute difference between working storage and its 5-year minimum value was included in the model. Contemporaneous working gas in storage could have been included in the model as an endogenous variable, but the process by which storage is determined is beyond the scope of the analysis. Therefore, a lagged value of working gas is used to reflect the information that suppliers had about market supply conditions, but could not directly influence.

[7] The results of the misspecification tests are reported in the Technical Appendix.

Including these exogenous variables resulted in improved model performance; however, some extreme residual values remained. Of these residuals in excess of 3 standard deviations from the fitted model, three seemed to result when extraordinary market and geopolitical events occurred. The first is August 1990, when Iraq invaded Kuwait. Oil prices shot up on news of the invasion, as fears that war in the Middle East could adversely affect oil markets. The second was February (and March) 1996 when extreme cold and uncertainty contributed to a large increase in the Henry Hub price in February , which was then followed by an offsetting decrease in March. The third is September 2001, when terrorist attacks on the World Trade Center and the Pentagon heightened uncertainty throughout the United States and its economy.

To model the effects of these extraordinary events, pulse dummy variables were created: the first was set equal to zero for all observations except August 1990; the second was set equal to zero for all observations except September 2001. To capture the effects of the transitory spike in natural gas markets in February 1996, a third dummy variable was created that was also equal to zero for all observations, but equal to 1 in February 1996 and -1 in March 1996. This was done because the spike in February 1996 was almost entirely offset by the decline in March 1996.

The pulse dummy variables were employed using considerable discretion and testing. The aim was to mitigate the effects of extreme observations, while estimating a model that was as parsimonious as possible. In each of these events, the extraordinary market conditions that prevailed at the time were the result of non-recurring, exogenous shocks rather than the normal evolution of market forces. The intent to capture their impact within the model warranted intervention dummy variables when including the observations in the estimation of the model. Moreover, these dummy variables were instrumental in resolving the leptokurtosis in the probability distribution of the residuals. The use of the intervention dummy variables ensured the multivariate normality of the probability distribution of the residuals, which permitted hypothesis testing on the results of the model, and ensured model stability.

The VAR(2) was re-estimated to include the effects of the exogenous variables and the dummy variables. The fitted model results and residuals demonstrate a much tighter fit with the data, with most observations of actual data lying within 2 standard deviations of the fitted model (Figure 7).. The residuals appear to be normally distributed, resulting from accounting for the presence of the extreme outliers using the pulse dummy variables (Figure 8). The model specification tests confirm that the re-estimated model does not have significant autocorrelation, heteroskedasticity, or non-normally distributed errors.

The Cointegration Analysis of the Vector Autoregression Model

In this section the Johansen procedure is applied to test for the presence of cointegration. The crux of the Johansen test is to examine the mathematical properties of the **Π** matrix in equation 7, which contains important information about the dynamic stability of the system. Intuitively, the **Π** matrix in equation 7 contains the expression relating the levels of the two endogenous variables, the Henry Hub and WTI spot prices. With two endogenous variables in the VAR, if the **Π** matrix has 1 linearly independent row, then crude oil and natural gas prices have a cointegrating relationship. In the specification with 2 endogenous variables there can be as many as 2 linearly independent equations, or no linearly independent equations.[8] Evaluating the number of linearly independent equations in **Π** is done by testing for the number of non-zero characteristic roots, or eigenvalues, of the **Π** matrix, which equals the number of linearly independent rows.[9]

Prior to conducting the test for cointegration, the possibility of deterministic trends in the data, in addition to stochastic trends, must be considered, because the asymptotic distribution of the test statistic used in the cointegration test is sensitive with respect to assumptions about deterministic trends in the data. From Figure 6, the expected values of the differenced prices series appear to be roughly equal to 0. This suggests that there is not a deterministic trend in the differenced series, and so including a trend variable outside of the cointegration relation in the estimation of equation 7 does not appear necessary. However, there may be a deterministic trend within the cointegration relationship, itself. In the lower panel of Figure 2, the Henry Hub series appears to be rising slightly faster than the adjusted WTI series, starting out consistently below the WTI series, until sometime in 1993, and thereafter exceeding the WTI series on a fairly regular basis. To test the possibility of a trend in the cointegrating relation, a trend variable will be restricted to the cointegrating equation. From equation 8, the model to be estimated when including the deterministic trend variable, becomes

(9) $\mathbf{\Delta p}_t=\mathbf{\Gamma}_1\mathbf{\Delta p}_{t-1} + \mathbf{\Gamma}_1\mathbf{\Delta p}_{t-2} + \boldsymbol{\acute{\alpha}}(\boldsymbol{\beta}'\mathbf{p}_{t-1} + \mu t) + \mathbf{Bx}_t + \boldsymbol{\acute{\varepsilon}_t},$

where

t is a trend variable.

Using the model specification in equation 9, the unrestricted VAR was estimated using ordinary least squares. The results of the Johansen cointegration test are presented in Table 1. The eigenvalues of the **Π** matrix are sorted from largest to smallest. The tests are conducted sequentially, first examing the possibility of no cointegrating relation against the alternative that there are two cointegrating relations, and then the null of one cointegrating relation against the possibility of more than 1 cointegrating equation. Essentially, this is a test of whether the eigenvalue is significantly different from zero. Both the Johansen trace test and max test support rejection of the null hypothesis that there are no cointegrating relations in the system. However, the Johansen tests are unable to reject the hypothesis that there is more than one cointegrating equation. These results

[8] If the rank of **Π** is equal to number endogenous variables then all of the original series are stationary; and if the rank of **Π** is zero, there are no cointegrating vectors.
[9] The number of linearly independent rows in a matrix is called the rank.

indicate that there is a single significant cointegration equation in the system, relating the long-run equilibrium between the Henry Hub natural gas and WTI crude oil spot prices.

Hypothesis Testing on the Estimated Coefficients of the Cointegrating Relationship

The next step in the analysis is to re-estimate the VAR, restricting it to incorporate only the significant cointegrating equation in the system. A useful, yet mathematically trivial, transformation is to normalize the **β'** vector. so, that one of the endogenous variables can be expressed in terms of the other and its parameter estimate. For example, expressing **β'**as **(**$\boldsymbol{\beta}_g$**' β**o**'**) yields:

(10) $\Delta \mathbf{p}_t = \boldsymbol{\Gamma}_1 \Delta \mathbf{p}_{t-1} + \boldsymbol{\Gamma}_1 \Delta \mathbf{p}_{t-2} + \boldsymbol{\acute{\alpha}}(\boldsymbol{\beta}_g(\mathbf{I}_2, \tilde{\boldsymbol{\beta}}_o')\ \mathbf{p}_{t-1} + \tilde{\mu}t) + \mathbf{B}\mathbf{x}_t + \boldsymbol{\acute{\varepsilon}}_t,$

where:

$\tilde{\boldsymbol{\beta}}_o' = (\beta g')^{-1} \boldsymbol{\beta} o'$; $\tilde{\mu} = (\beta g')^{-1} \mu o$

$\beta g' <> \mathbf{0}$

This transformation does not alter the equation or the likelihood function, and so it is not necessary to perform any perform any statistical testing on the restriction on this restriction. The normalization enhances the economic interpretability of the long-run equilibrium expression by expressing the cointegrating relation as a deviation from the natural gas price.

The estimated **β** coefficients and associated standard errors of the cointegrating relation for the Henry Hub natural gas price equation are present in Table 2, along with the **ά** coefficients and associated standard errors. Owing to the normalization in the preceding paragraph, the **β** coefficient for the natural gas price is 1 and its associated standard error is 0. The cointegrating relation may be interpreted as expressing the Henry Hub price as a function of the WTI price and a trend term, so that the Henry Hub price is equal to 81 percent of the crude oil price plus a trend term of about 0.51 percent.

Likelihood ratio tests on the coefficients are performed by imposing the hypothetical restrictions and comparing the impacts on the characteristic roots. Each coefficient is tested against the null hypothesis that they are equal to zero. The results in Table 2 indicate that the **β** coefficients are statistically significant, strongly rejecting the null hypothesis. This indicates that the cointegrating relationship will not benefit from over-identifying the parameters and the relationship may be written:

(11) $ECM_t = p_{g,t} - 0.83 p_{o,t} + 0.005t$

where ECM_t is defined as the equilibrium correction mechanism.

The cointegrating relation may be interpreted depicting departures from a long-run equilibrium. In Figure 9, the long-run equilibrium that is expected to prevail is given by the zero-line where there was an equilibrium in a long-run sense between natural gas prices and oil prices. In this way, natural gas and crude oil prices vary together around the fixed long-run equilibrium at the zero-line. When a disequilibrium results from a shock to crude oil or natural gas prices, it is followed by a return to the long-run equilibrium. For example, the peak in February 1996 shows the impact of the severe freezing weather nation-wide on natural gas prices as the Henry Hub price climbed to $4.42 per MMBtu compared with the preceding month's price of $2.92 per MMBtu. Similarly, the peak in December 2000 also shows the impact of severe freezing weather as Henry Hub prices rose from $5.52 MMTU in November to $8.90 MMBtu in December. In each case these departures from the long-run equilibrium were eventually followed by return to the long-run equilibrium.

Next, hypothesis testing is done on the **ά** parameters to ascertain whether one of the endogenous variables in the system influences the other variable but is not influenced by it. This is known as a test of long-run weak exogeneity. From Table 2, it is apparent that the **ά**-coefficient corresponding to the WTI price is small relative to its standard deviation while the **ά**-coefficient corresponding to the Henry Hub price is large relative to its standard deviation. This suggests that the cointegrating relation does not have a significant influence on the evolution of the WTI price, but is highly significant for the Henry Hub price equation. Formally, tests of weak exogeneity were also done using the likelihood ratio. The results are presented in Table 3, providing statistically significant evidence that oil prices are weakly exogenous to natural gas prices.

With the finding of the weak exogeneity of crude oil prices it becomes possible to re-estimate the system of equations as a partial system with one equation for natural gas prices. Weak exogeneity implies that the long-run equilibrium relation does not have explanatory power for the crude oil price, and so the crude oil price equation may be dropped from the model, and the natural gas price equation can be conditioned on the contemporaneous crude oil price. The final specification and estimation of the model may be expressed as:

$$(12) \quad \Delta p_{g,t} = -0.50 + 0.30\Delta p_{o,t} + 0.10\Delta p_{g,t-1} - 0.21ECM_{t-1} + Bx_t + \acute{\varepsilon}_t,$$

Estimating the model in this fashion has the benefit of increased simplicity in more ways than one. Because crude oil prices are now treated as an exogenous variable, the impulse dummy variable for August 1990 may now be dropped from the model. Tests of exclusion were conducted on each of the remaining deterministic variables, including the impulse dummy variables. Outlier dummy variables for February 1997 and September 2001, as well as the transitory dummy variable for February/March 1996 were necessary to ensure multivariate normality and homoskedasticity. However, several other outlier dummy variables were found to have significant explanatory power, and so they were included in the model to ensure the best fit to the data. The complete model results including the exogenous and deterministic variables are presented in Table 4.

Implications of the Model and Model Dynamics

Equation 12 indicates that a 10 percent increase in contemporaneous crude oil prices will lead to an increase of three percent in natural gas prices in the first month. The parameter of the lagged natural gas price term has a similar interpretation with a 10 percent increase in the lagged natural gas price implying 1 percent in crease in the contemporaneous price. The speed of adjustment parameter, 0.21, indicates that approximately 21 percent of a difference from the long-run equilibrium will be recovered in the following period.

A key finding of the model is that adjustments to differences from the long-run equilibrium may be used to explained short-run movements in price. This makes the model quite flexible in capturing the effects of large stochastic shocks that may cause disequilibria. This also underscores the finding that natural gas and crude oil prices have had a stable long-run relationship despite periods when a large exogenous spike in either crude oil or natural gas prices may have produced the appearance that these two prices had decoupled.

A couple of simple examples will illustrate how natural gas prices respond to changes in the crude oil price in the short-run, holding all else constant. Two cases are considered: a permanent increase in the crude oil price, and second, a transitory increase in the crude oil price (Table 5).

The first case is presented in the upper panel: Crude oil prices rise by 20 percent, and remain unchanged thereafter. The immediate response of natural gas prices is to climb about 5 percent in the same period as the oil price shock occurs. In each successive period following the initial shock, gas prices continue to climb, but at decreasing rate as gas prices adjust to the earlier disruption in the long-run equilibrium. Gas prices climb about 10 percent within the first two months following the shock, and about 15 percent within 12 months.   The oil price shock is not passed through on a one-to-one basis to natural gas prices.

The second case is presented in the lower panel: A transitory spike in crude oil prices occurs such that the crude oil price increases 20 percent in the period 0, followed by an offsetting decrease in the period 1. As in the preceding case, natural gas prices respond to the contemporaneous increase in crude oil prices by climbing about 5 percent. In the period following the initial price spike, the natural gas price would have continued to climb to restore the long-run equilibrium. However, it is hit with a second contemporaneous oil price shock—this time a downward movement in period 1 of about 17 percent in oil prices, which returns the oil price to its original level. The second shock causes a short-term impact that more than offset the long-run adjustment. This leads the natural gas price to decline about 3 percent in the period 1 to only 2 percent above its initial starting level. 2 months following the initial oil price shock (period 2), the natural gas price is about 1 percent above the price level in the initial period. In the ensuing months, the natural gas price continues to decline. At 12 months past the initial shock, it is about 1 percent below its initial level.

Summary:

Economic theory suggests that there is a relation between natural gas and oil prices, but the influence of an increase in oil prices may conflict in its effects on natural gas supply and therefore prices. Production of natural gas may increase as a co-product of oil, or may decrease as a result of higher-cost productive resources. While the net effect of an increase in oil prices on natural gas supply may be ambiguous, the effect on natural gas demand is clear, resulting in a positive relation between oil and gas prices. Given the relative inelasticity of natural gas supply in the short-term owing to factors such as a 12-18 month lag in the production response to drilling changes, it appears that the effect of oil prices on natural gas demand is dominant in the short run. These issues were explored empirically using cointegration analysis and a simple vector error correction model.

The analysis supports the presence of a cointegrating relationship between the crude oil and natural gas price time series, providing significant statistical evidence that WTI crude oil and Henry Hub natural gas prices have a long-run equilibrium relationship. A VECM of crude oil and natural gas prices was estimated, and facilitated the analysis of the long-run equilibrium, and short-run adjustments in prices. The estimation of the model resulted in identifying evidence of a stable relationship between natural gas and crude oil prices. The statistical evidence also supported the a priori expectation that while oil prices may influence the natural gas price, the impact of natural gas prices on the oil price is negligible. With crude oil prices weakly exogenous to natural gas prices, the short run response of the natural gas price to contemporaneous changes in the oil price was found to be statistically significant. Another significant finding of the model was that natural gas prices appear to be growing at slightly faster rate than crude oil prices, suggesting a narrowing of the gap between the two over time. Finally, natural gas and crude oil prices historically have had a stable relationship, despite periods where they may have appeared to decouple. Future studies should explore applying the findings the stable long-run relationship between oil and gas prices in both long-run and short-run forecasting.

Figure 1: Henry Hub and WestTexas Intermediate Prices (1989-2005)

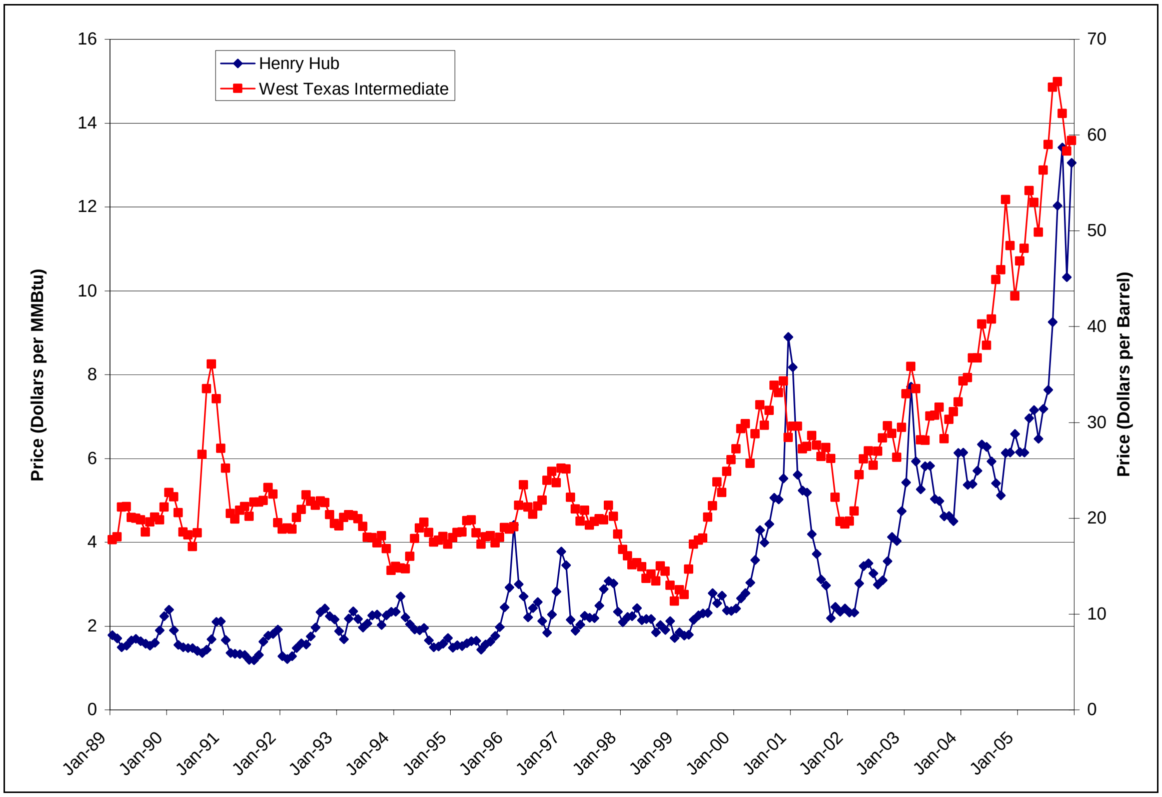

Figure 2: Henry Hub and West Texas Intermediate Prices and Mean-Adjusted West Texas Intermediate in Logarithms (1989-2005)

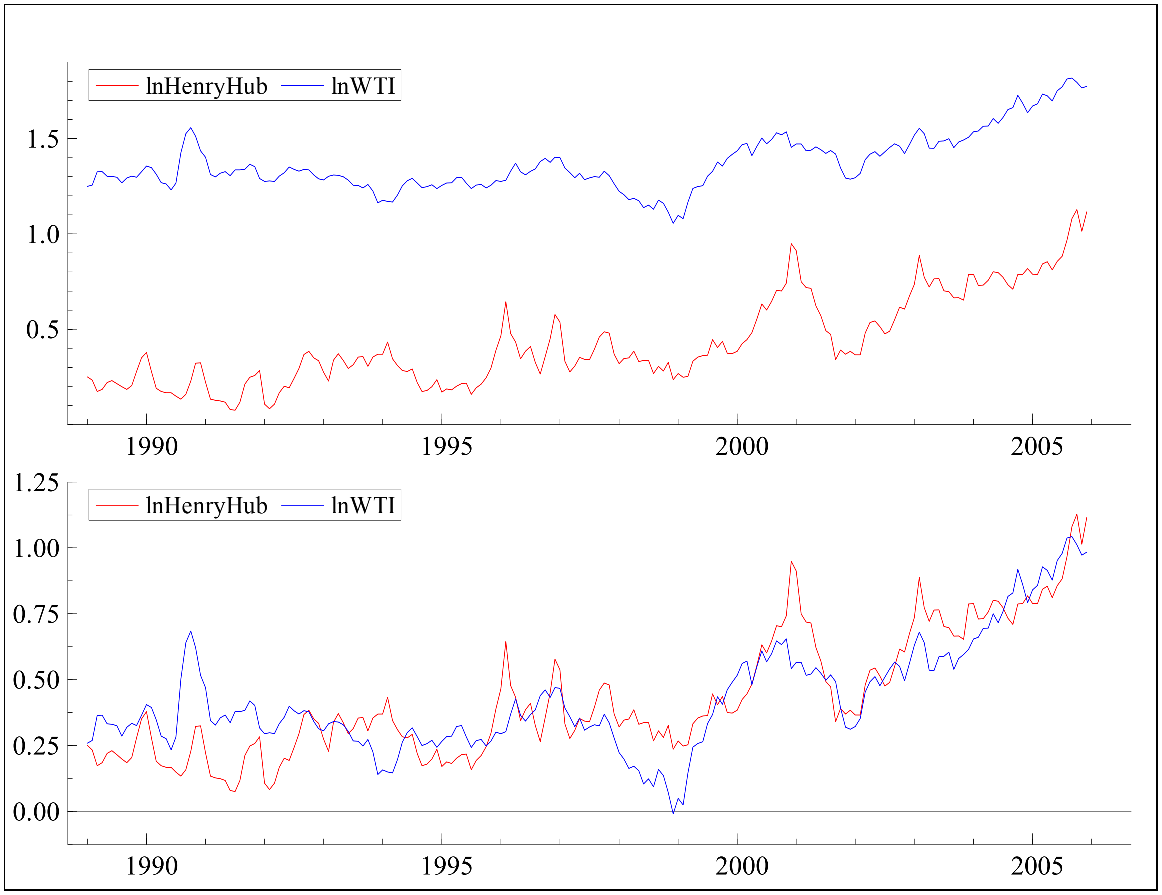

Figure 3: Henry Hub and West Texas Intermediate Prices Histogram and Autocorrelogram Functions

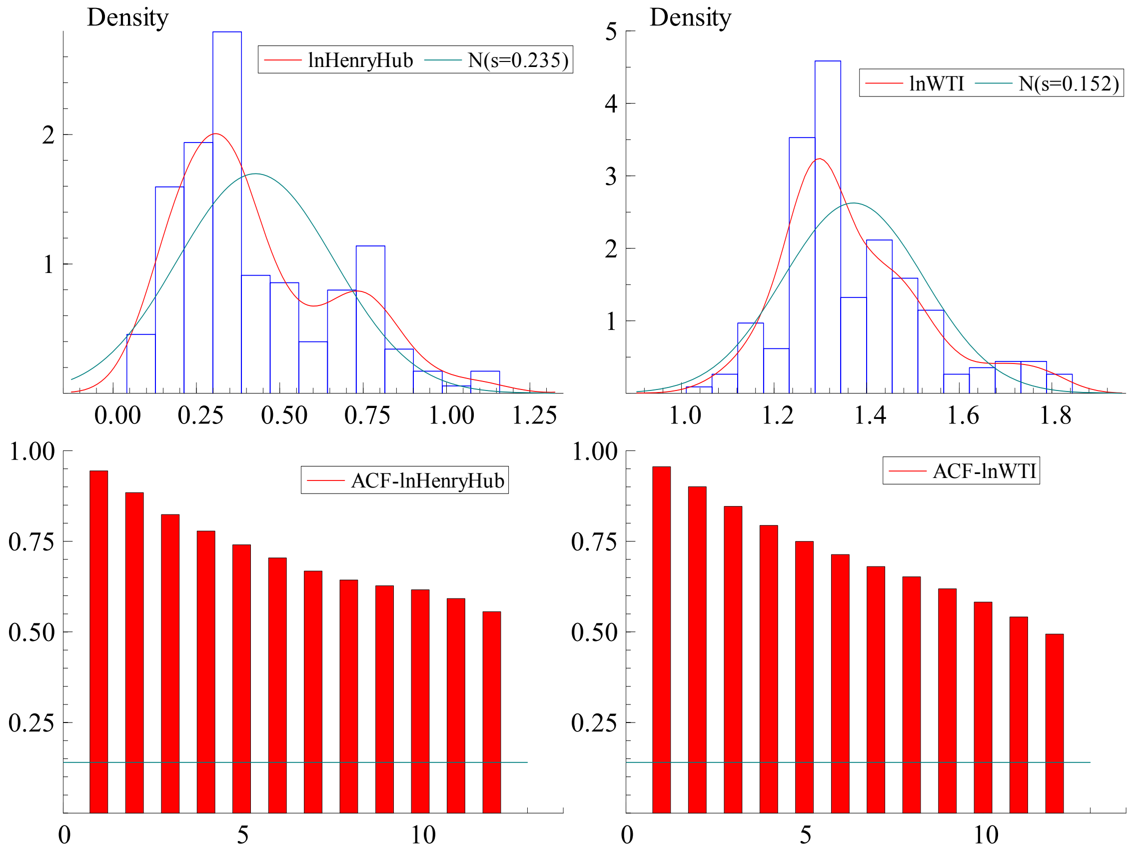

Figure 4: First Differences of Henry Hub and West Texas Intermediate Prices Histogram and Autocorrelogram Functions

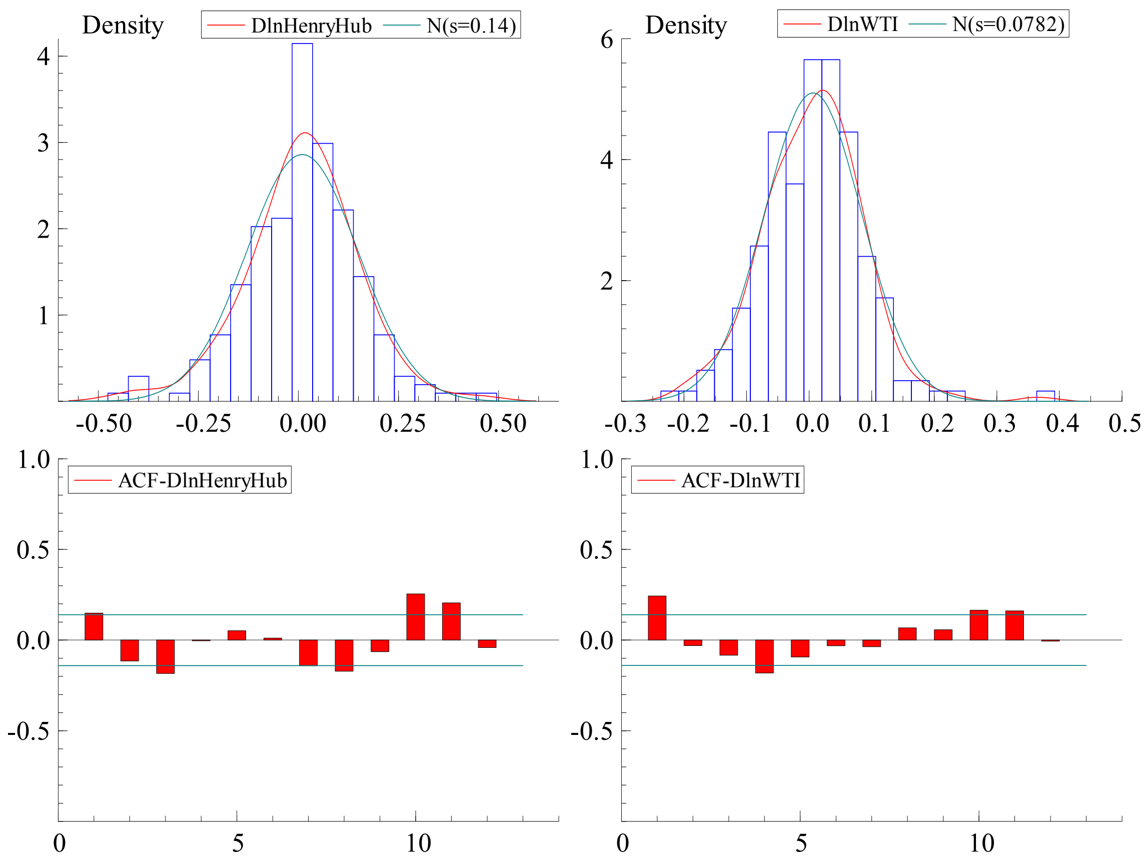

Figure 5: Henry Hub and West Texas Intermediate Prices with Fitted Values and and Residuals from a VAR(2)

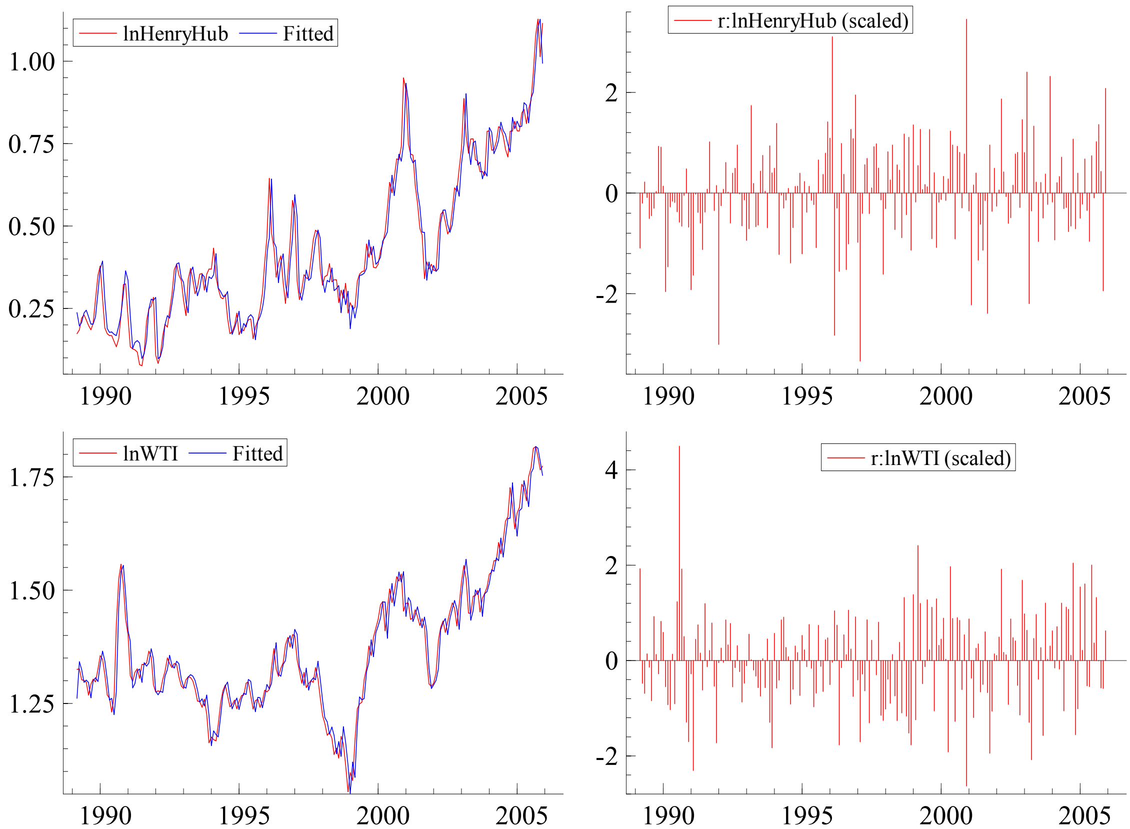

Figure 6: Autocorrelogram and Empirical Density of VAR(2) Residuals

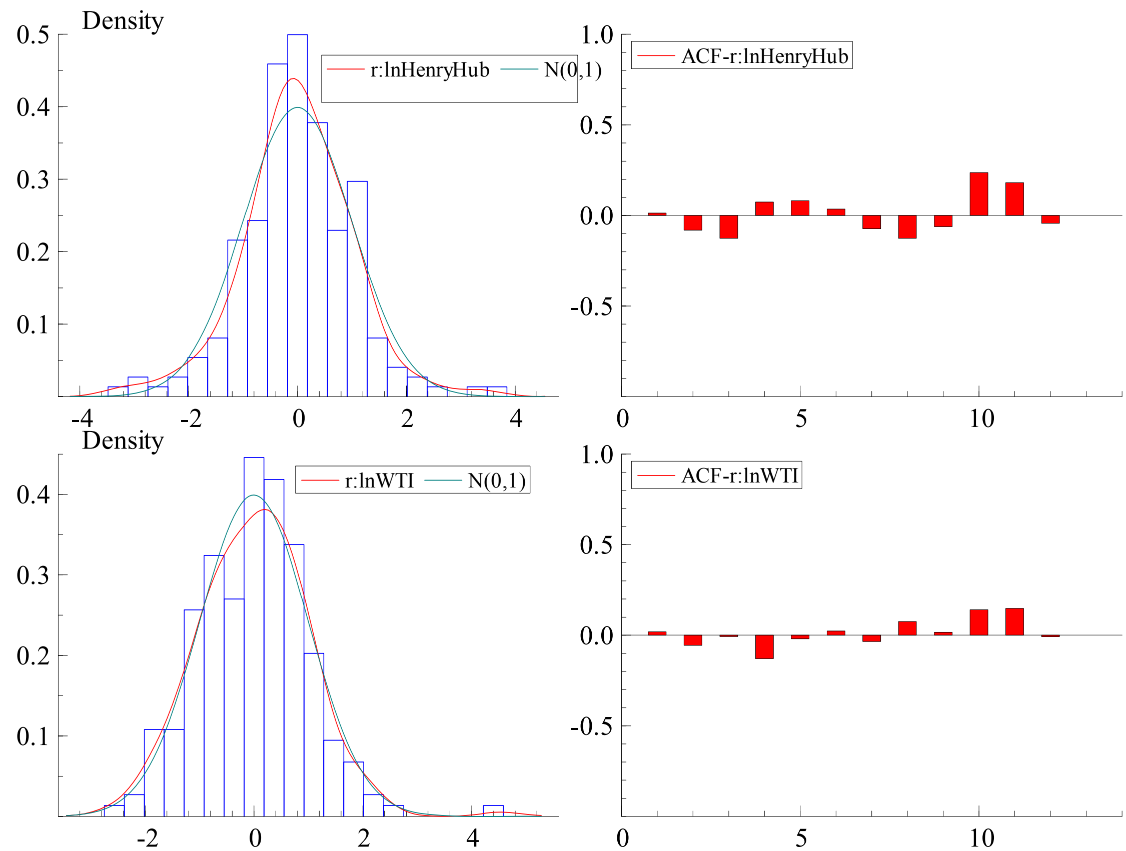

Figure 7: Henry Hub and West Texas Intermediate Prices with Fitted Values and and Residuals from Unrestricted VAR(2) Complete Model

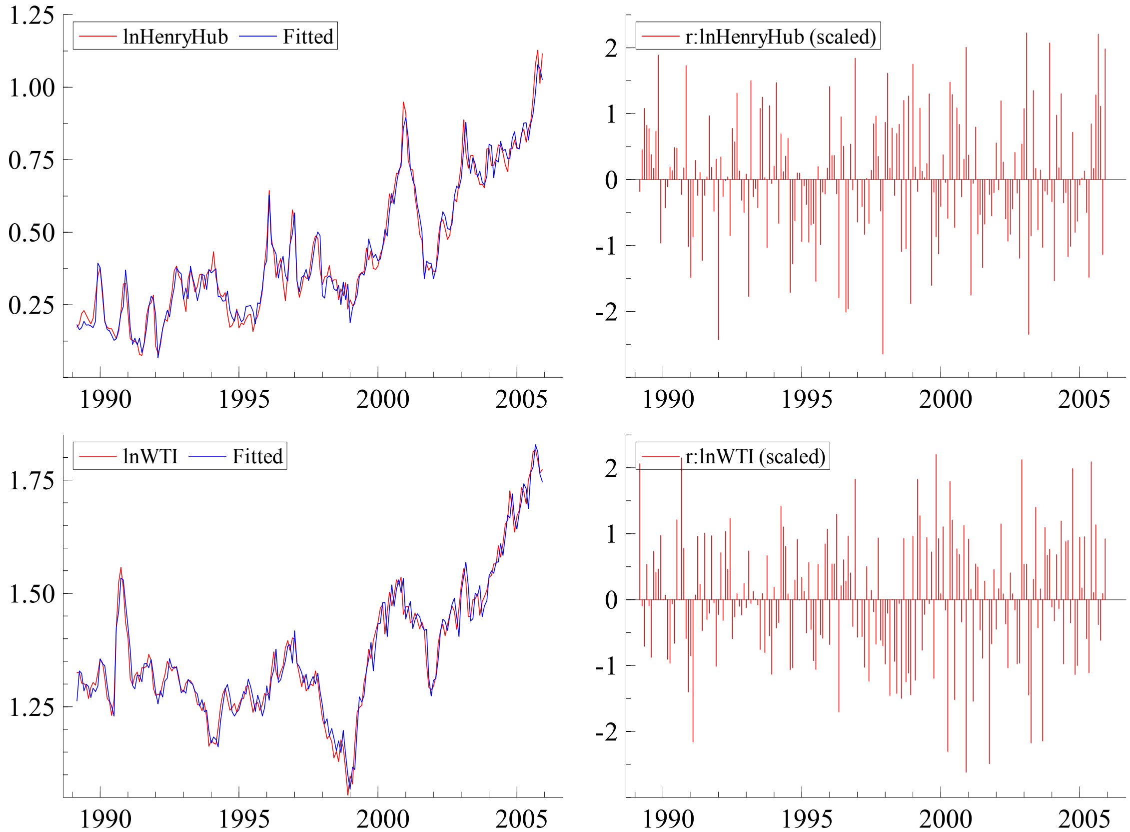

Figure 8: Autocorrelogram and Empirical Density of Unrestricted VAR(2) Complete Model Residuals

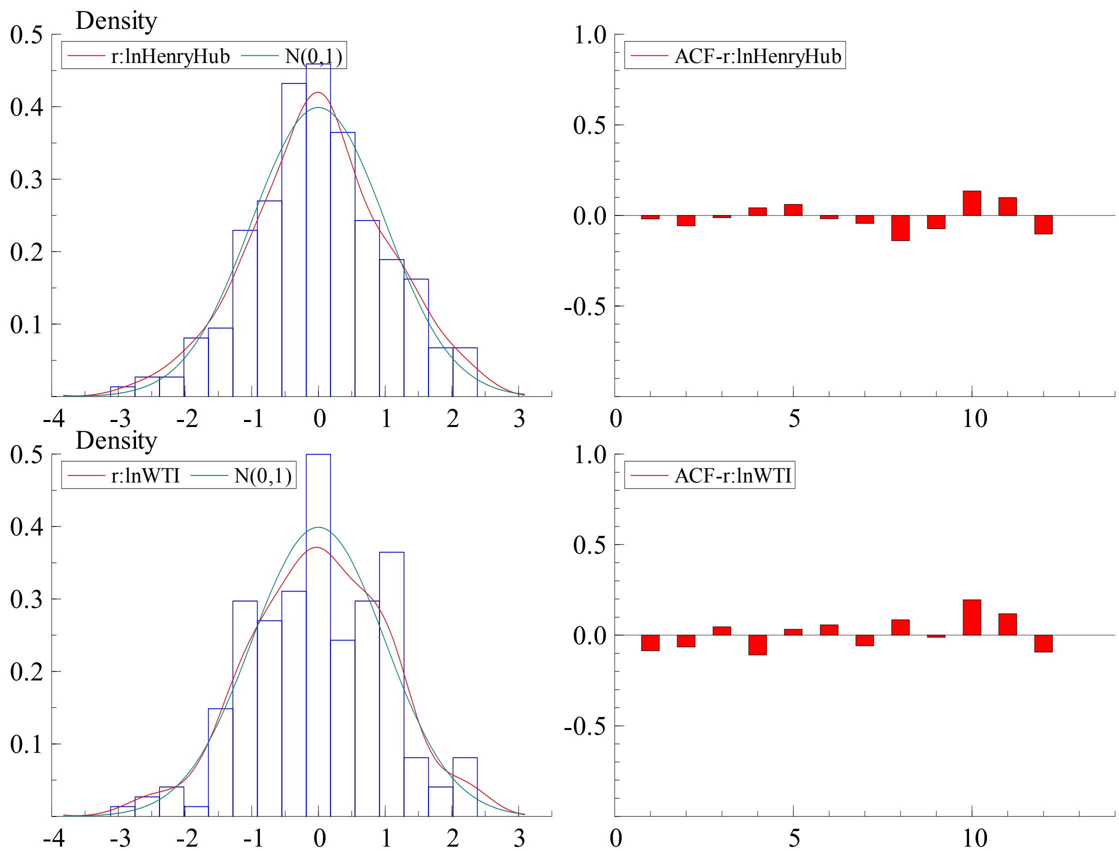

Figure 9: The Estimated Cointegrating Vector Between Henry Hub and WTI Prices (1989-2005)

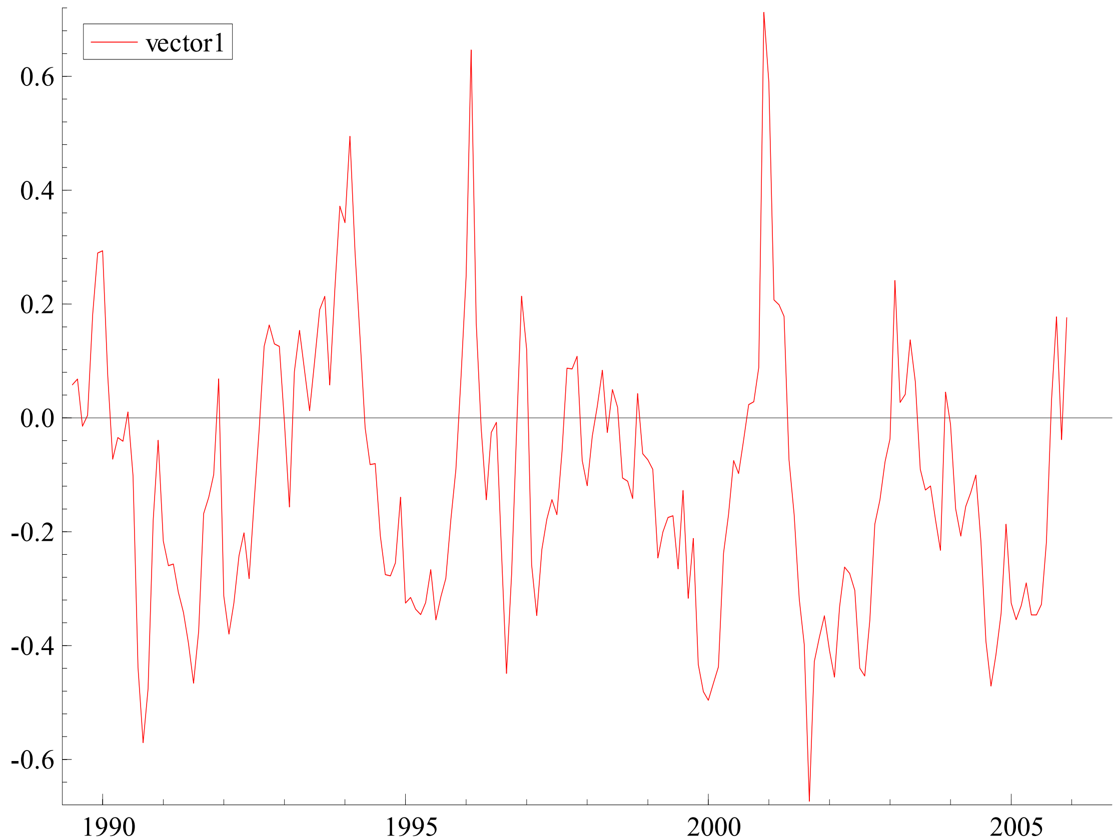

Table 5. Effects of a Permanent and Transitory Change in Crude Oil Prices

Effects of a Permanent Increase in the Crude Oil Price

Period	Percentage Change WTI Crude Oil Price	Cumulative Percentage Change Henry Hub Price
0	20.0%	5.3%
1	0.0%	7.8%
2	0.0%	9.6%
12	0.0%	14.8%

Effects of a Transitory Increase in the Crude Oil Price

Period	Percentage Change WTI Crude Oil Price	Cumulative Percentage Change Henry Hub Price
0	20.0%	5.3%
1	-16.7%	2.2%
2	0.0%	1.2%
12	0.0%	-1.0%